

Report: Sonar Data Analysis and Mine/Rock Prediction

Date: 11-07-2025

This notebook demonstrates the process of loading, analyzing, and building a classification model to predict whether an object is a mine or a rock based on sonar data.

1. Data Loading and Exploration:

- The dataset was loaded from a Google Drive link into a pandas DataFrame.
- Initial exploration using `df.head()`, `df.shape`, `df.describe().T`, and `df.info()` provided insights into the data structure, size, descriptive statistics, and data types.
- The target variable (column 60) was examined using `value_counts()` and `groupby().mean()` to understand the distribution of 'Mine' (M) and 'Rock' (R) classes and the average feature values for each class.

2. Data Preparation:

- The data was separated into features (X) and the target variable (y).
- The dataset was split into training and testing sets using `train_test_split` with a test size of 10%, a random state for reproducibility, and stratification to maintain the proportion of classes in both sets.

3. Model Training and Evaluation:

- A Logistic Regression model was initialized and trained on the training data.
- **Training data accuracy:** 0.8342
- **Test data accuracy:** 0.7619

4. Prediction:

- A sample input from the training data was selected and reshaped to be compatible with the model's predict method.
- The model predicted whether the sample input represented a 'Rock' or a 'Mine'.

Conclusion:

The Logistic Regression model achieved an accuracy of 0.8342 on the training data and 0.7619 on the unseen test data. This indicates that the model has learned to distinguish between 'Mine' and 'Rock' based on the sonar data, although there is a slight drop in accuracy on the test set, which is expected. Further model tuning or exploration of other classification algorithms could potentially improve the test accuracy.